

# К физической теории мышления: теория нейронных моделирующих полей

Перловский Л.И.

Гарвардский университет, leonid@deas.harvard.edu

**Аннотация.** В статье сделан обзор подходов к моделированию мышления в таких направлениях исследований как искусственный интеллект, распознавание образов, нейронные сети. Дается математическая формулировка нейронных моделирующих полей, объединяющая механизмы мышления в систему, соответствующую данным психологии и нейробиологии. Подчеркивается роль инстинкта к знанию в активном процессе познания окружающего мира.

**Ключевые слова:** мышление, физика, эмоции, концепции, сознание, инстинкт к знанию.

## 1. Физика и мышление

Методы нейронных сетей в настоящее время активно используются для решения разнообразных практических задач. Однако известные нейронные сети слишком просты для объяснения человеческого мышления. В этой статье мы попытаемся не просто улучшить методы для решения практических задач, но также сделать шаг к тому, чтобы объяснить мышление, как физический процесс, так же как Ньютон объяснил движение планет. Но возможна ли физическая теория мышления и в чём могла бы быть её суть? Предполагается, что, говоря о мышлении и мозге, мы обсуждаем одну и ту же физическую систему на разных уровнях описания. Эта ситуация для физики не нова, например, термодинамика и статистическая физика связаны подобным образом. Физическое описание на микроскопическом уровне (атомы, молекулы, элементарные частицы, струны) предпочтительно не всегда, часто феноменологическое описание более эффективно. Эйнштейн однажды заметил, что термодинамика его привлекает именно тем, что определена на феноменологическом уровне. Физическая теория отличается поиском основных законов, нескольких универсальных принципов, из которых математически следует описание экспериментальных данных в широкой области.

Проблема мышления обсуждалась в течение тысячелетий, от философов Древней Греции до современных математиков и исследователей в области когнитивных наук. И всё же терминология в этой области ещё не установилась. Слова: *мышление, мысль, воображение, эмоции, концепции, концепты* используются различным образом в повседневной жизни, также они не были чётко определены в когнитивной науке и являются предметом дискуссий [1]. Мы считаем, что стандартизированные определения уместны лишь после того, как теория создана (например, «сила» была определена ньютоновскими законами после тысяч лет менее точного использования). Поскольку теория мышления ещё соз-

даётся, в данной статье придерживаемся следующего порядка словоупотребления. Предлагаемые термины должны: (1) соответствовать современным дискуссиям в научной и математической среде, (2) соответствовать тысячелетним философским обсуждениям и принятому употреблению в культуре, (3) быть чёткими и описываемыми математически, и (4) отклонения от этих принципов и противоречия в употреблении терминов должны отмечаться и обсуждаться. В соответствии с определением в словаре [2], которое мы принимаем лишь как отправную точку, мышление включает сознательные и бессознательные процессы, основными среди которых являются не только мысль как концептуализация, но также восприятие, эмоция, воля, память, воображение; мышление формируется в мозге. Мнения о том, какие математические методы подходят для описания мышления, сильно различаются. Основатели искусственного интеллекта, включая Аллана Ньюэлла и Марвина Минского, считали, что для этой цели адекватна формальная логика [3,4] и никакие специальные новые математические методы не требуются. Противоположная точка зрения развивалась Брайаном Джозефсоном и Роджером Пенроузом, считавшими, что мышление не может быть понято на основе современной физики; для объяснения мышления потребуются новые неизвестные ещё физические законы и явления [5,6]. Рассматривались квантовые процессы, которые могут происходить в нейронах [5-8]. Некоторые авторы, рассматривая процессы классической физики в нейронах, считают, что небольшое число основных принципов нейронной организации играют фундаментальную роль, определяя основные законы мышления, и необходимо найти математическое описание этих основных законов. Этому мнению придерживаются Стефан Гроссберг, предложивший, что основные законы включают резонансное совпадение афферентных и эфферентных сигналов (идущих сверху-вниз и снизу-вверх) [9] и эмоциональную оценку концептуального содержания [10], Лотфи Заде, предложивший теорию гранул [11], Алекс Мейстель, разрабатывающий иерархическую многомасштабную организацию мышления [12]; Джералд Эдельман, предложивший отбор нейронных групп [13]; и автор данной статьи, предложивший инстинкт к знанию, эстетические эмоции и динамическую логику [8,14-16].

## 2. Логика и мышление

Длительное время считалось, что интеллект эквивалентен концептуальному пониманию и рассуждению. Часть этой «веры» состояла в том, что мышление основано на логике. Основатели искусственного интеллекта в 1950-х и 1960-х годах верили, что, основываясь на правилах логики, они вскоре создадут компьютеры, чей интеллект намного превзойдёт человеческий.

Начало этой истории обычно относят к Аристотелю, сформулировавшему правила логики [17]. Однако Аристотель не считал, что мышление работает логически; он открыл логику, как лучший способ аргументации, а не как теорию мышления. Например, в книге «Риторика для Александра» Аристотель перечисляет десятки тем, на которые Александру Македонскому приходится произносить политические речи [18]. На каждую тему Аристотель приводит два противоположных мнения (например, заключать мир или объявлять войну, использовать пытки при допросе или нет...). Для каждого из противоположных

мнений Аристотель даёт свои логические аргументы. Ясно, что логика, по Аристотелю, – это инструмент для выражения ранее принятых решений, а не механизм мышления. Логика может лишь быть использована для дедукции из основных принципов, но не способна указать, в чём эти принципы должны состоять. В последующие столетия тонкости мысли Аристотеля не всегда верно понимались. С развитием науки идея, что интеллект эквивалентен логике, стала захватывать умы. В XIX веке математики обратились к логике. Джордж Буль заметил противоречие в теории Аристотеля. Начиная от Аристотеля, логика основывалась на законе исключённого среднего (или исключённого третьего): каждое утверждение либо истинно, либо ложно, любая средняя альтернатива исключается [19]. Но Аристотель также подчёркивал, что логические утверждения не должны формулироваться слишком точно (например, не следует определять меру пшеницы – мешок – с точностью до одного зерна), что язык подразумевает необходимую неопределённость.

Буль считал, что следует исправить противоречие между точностью закона исключённого среднего и нечёткостью языка. Родилась новая область математики, *формальная логика*, которой занимались Джордж Буль, Готтлоб Фреге, Георг Кантор, Бертран Рассел, Давид Гильберт, Курт Гёдель. Логика «выбросили» нечёткость языка и основали формальную математическую логику на законе исключённого среднего. Гильберт разрабатывал направление названное формализмом, отрицающее интуицию, и хотел определить объекты науки формально, в терминах логических аксиом. Он был уверен, что его логическая теория также описывает механизмы мышления: «Основная идея моей теории доказательств – это не что иное как описание нашей деятельности понимания, это протокол правил, согласно которым наше мышление в действительности происходит» [20]. В 1900 г. он сформулировал знаменитую Entscheidungsproblem: определить систему логических правил, достаточную для доказательства всех прошлых и будущих математических теорем. Это подразумевало формализацию научного творчества и всего человеческого мышления.

Почти сразу, после того как Гильберт сформулировал свою программу формализации, обнаружилась первая дыра. В 1902 г. Рассел показал несостоятельность формальной процедуры, определив множество  $R$  следующим образом:  *$R$  есть множество всех множеств, которые не включают самих себя*. Включено ли  $R$  в  $R$ ? Если нет, то  $R$  должно быть включено в  $R$  по определению, но если  $R$  включено в  $R$ , это противоречит определению. В любом случае получаем противоречие. Этот пример получил название парадокс Рассела. В последовавшие 25 лет математики пытались построить самосогласованную математическую логику, свободную от подобных парадоксов. Но в 1931 г. Гёдель доказал, что это невозможно [21], формальная логика оказалась противоречивой, не самосогласованной [22].

Пенроуз посвятил книгу [5] вопросу о мышлении в свете теории Гёделя. Он пришёл к выводу, что мышление не может быть объяснено как вычислительная процедура (то есть не может быть описано существующей математикой и не может быть реализовано ни в компьютере, ни в какой системе классической физики). Разумеется, Пенроуз встретил много возражений среди коллег. Книга [5] разбирает эти возражения, и Пенроуз подчёркивает, что рассмотрел все возражения. Настоящая статья, однако, даёт новое возражение: Логика не

является ни фундаментальным механизмом мышления, ни его основной. Способность к логическому мышлению возникает из нелогических механизмов мышления.

### 3. Обзор математических теорий интеллекта. Логика и сложность

Простое восприятие вовлекает сигналы из сенсорных органов и внутренние представления мышления (память) об объектах. В процессе восприятия мышление ассоциирует подмножества сигналов, соответствующих объектам, с представлениями об этих объектах. Далее активируются нейронные сигналы, ведущие к мысленным и поведенческим реакциям, являющимся частью понимания объекта или ситуации.

Математическое описание этого *первого шага*, казалось бы, простого процесса ассоциация-распознавание-понимание оказалось далеко не простым делом, и это связано с понятием комбинаторной сложности (КС) [23]. Это понятие подчёркивает многочисленность комбинаций элементов в сложной системе; например, восприятие ситуации обычно требует одновременного распознавания её многочисленных элементов, которые могут встретиться в различных комбинациях. Например, рассмотрим 100 элементов (не слишком большое число); число комбинаций из 100 элементов,  $100^{100}$  превышает число всех элементарных частиц во Вселенной за всё время её существования. Никакой компьютер не будет в состоянии провести столько вычислений.

Эта проблема впервые была замечена в задачах распознавания образов в 1960-х годах и получила название «проклятие размерности» [24]. Казалось, что адаптивные самообучающиеся алгоритмы и нейронные сети могли бы обучиться решению любой задачи «сами по себе», они лишь нуждались в достаточном числе обучающих примеров. Последовавшие тридцать лет создания адаптивных самообучающихся алгоритмов и нейронных сетей привели к выводу, что требуемое число обучающих примеров часто оказывалось комбинаторно большим. Таким образом, самообучающиеся методы встретились с *КС обучения*.

В 1970-х годах для решения проблемы КС были предложены *системы правил* [25,26]. Первоначальная идея состояла в том, что знания, необходимые для решения задач определённого круга, будут сформулированы в виде логических правил, и необходимость в обучении отпадёт. Однако в непрерывно изменяющемся мире количество необходимых правил росло; правила оказались зависимы от других правил и необходимо рассматривать комбинации правил. Системы правил встретились с *КС правил*.

Начиная с 1980-х годов, были предложены *системы моделей*. Они использовали модели, зависящие от адаптивных параметров. Идея состояла в том, чтобы соединить преимущества систем правил с адаптивностью и обучаемостью. Знания предполагалось вложить в модели, а изменяющимся аспектам конкретных ситуаций система должна была обучаться сама, подгоняя параметры моделей [27,28]. Подгонка моделей к данным требовала идентифицировать подмножества данных, соответствующих различным моделям. Количество подмножеств, однако, комбинаторно велико. Популярный алгоритм для одновре-

менной подгонки многих моделей к данным, «проверка многих гипотез» [29], известен тем, что сталкивается с КС вычислений. Системы моделей встретились с *КС вычислений*.

Последовавшие исследования обнаружили связь КС с логикой в различных алгоритмах [23]. Формальная логика использует «закон исключённого среднего», который исключает возможность ввести меру близости логических высказываний. Поэтому, алгоритмы, использующие логику, рассматривают каждое даже небольшое изменение в данных или моделях, как новое высказывание (гипотезу); более того, комбинации данных рассматриваются как новые высказывания. Это приводит к КС. Более того, КС алгоритмов, основанных на логике, связана с гёделевской теорией: проявляется несогласованность логики в конечных системах [30]. Многозначные логики и нечеткая логика были предложены для того, чтобы преодолеть ограничения, связанные с законом исключённого третьего [31]. Однако, математика многозначных логик не отличается в принципе от формальной логики, «исключенное третье» заменяется на «исключенное  $n+1$ ». Нечеткая логика встретилась со сложностью степени нечеткости, если берётся слишком большая нечеткость, решение не достигает необходимой точности: если нечеткость слишком мала, метод приближается к формальной логике. Сложные системы требуют различных степеней нечеткости в различных частях системы, изменяющихся в процессе работы системы. Поиск подходящих степеней нечеткости для различных комбинаций элементов системы вновь приводит к КС.

Возможна ли логика после Гёделя? Недавний обзор этой области [32], показывает, что логика после Гёделя усложнилась, стала «менее логичной», чем предполагалось основателями искусственного интеллекта, и оказывается напрямую связанной с проблемой КС. Повторю, Пенроуз считал, что из теории Гёделя следует невычислимость мыслительных процессов и необходимость новой физической теории [5]. Противоположная точка зрения, высказываемая в данной статье, состоит в том, что из невычислимости логики не следует невычислимость мышления. Логика не является основным механизмом мышления.

#### 4. Структуры мышления

В 1950-60-х годах исследователи в области искусственного интеллекта наивно верили, что вскоре они создадут компьютеры, превосходящие человеческий интеллект, и что математика логики окажется достаточной для этой цели. Однако, как мы обсудили, логика не объясняет мышления, но мышление работает. Обратимся к механизмам мышления, основываясь на представлениях, сложившихся в психологии, философии, когнитивной науке и нейробиологии. Основные механизмы мышления включают инстинкты, концепции, эмоции, поведение.

Инстинкты – это врождённые способности, не обучаемые, обычно адаптивные. В человеке и высших животных инстинкты связаны с эмоциями.

Слово концепты или концепции мы используем для обозначения общего смысла таких понятий как концепция, идея, понимание, мысль, представление, понятие. Смысл этих слов состоит в обозначении категории, группы, или класса предметов, явлений, событий или отношений. Концепции абстрактны в

том смысле, что они не различают индивидуальные явления. Подчёркивая это свойство, средневековые философы использовали слово «универсалии». Платон и Аристотель называли их идеями или формами и считали их основой для способности к пониманию мира. Подобным образом Кант считал, что они составляют содержание чистого разума, механизма способности к пониманию. По Юнгу обучение сознательным концепциям основано на врождённых бессознательных структурах психики, архетипах. Современная наука обычно связывает механизмы концепций с внутренними представлениями-понятиями об объектах и об отношениях между ними.

Эмоциями называют внутренние состояния, связанные с чувствами, а также преувеличенно-экспрессивное поведение, используемое для коммуникаций чувств. Многие исследователи (например, Левайн и Ливен [33]) подчёркивают, что в каждом решении обычно присутствуют в той или иной мере и эмоции и логика.

Важная роль эмоций в мышлении разбиралась многими авторами с различных точек зрения: в философии (Ренэ Декарт [34], Иммануил Кант [35], Жанн Поль Сартр [36]); в аналитической психологии (Карл Юнг [37]); в психологии и нейронных механизмах (Стефан Гроссберг и Даниэл Левайн [10], Эндрю Ортони [30], Джозеф ЛеДу [39]); в философской лингвистике (П. Гриффис [40]); в нейрофизиологии (Антонио Дамазио [41]); в адаптации, обучении и познании (Л.И. Перловский [16,42]).

Декарт попытался дать научное описание эмоций («страстей»). Он рационализировал эмоции, объяснил их как объекты и соотнёс с физиологическими процессами. По Канту эмоции тесно связаны с суждениями о том, какие конкретные восприятия и ситуации соответствуют каким общим концепциям-понятиям (и наоборот). Эмоции оказываются неотъемлемой частью мышления. Способность к эмоциям-суждениям – основание для всех высших познавательных способностей, включая способность к прекрасному и возвышенному. Сартр приравнивал эмоции в значительной мере к бессознательному содержанию психики; сегодня это не кажется адекватным. Юнг проанализировал сознательные и бессознательные аспекты эмоций. Он указал на недифференцированный статус примитивных психологических состояний, в которых эмоции, концепции и поведение слиты, и на их роль в повседневной жизни и в психозах. Он также подчеркнул рациональный аспект сознательных дифференцированных эмоций. Ортони объяснял эмоции в терминах концептуально-понятийных представлений и подчёркивал роль абдуктивной логики, как механизма для определения эмоций других людей. ЛеДу анализировал нейронные структуры и пути распространения нейронных сигналов, участвующих в эмоциональных процессах, в особенности, при страхе. Гриффис рассматривает основные эмоции и их эволюционное развитие в социальном взаимодействии. Дамазио считает, что эмоции – это в основном телесные ощущения и восприятие эмоций в мозгу вовлекает «телесные точки». Гроссберг и Левайн считают, что эмоции – это нейронные сигналы, связывающие инстинктивные и концептуальные мозговые центры. В процессах восприятия и познания эмоции оценивают насколько концепции-модели объектов и ситуаций удовлетворяют или не удовлетворяют инстинктивные потребности. Математическое описание в следующей секции близко связано с идеями Канта, Юнга, Гроссберга и Левайна.

Объяснение основных механизмов мышления, повторю, не требует мистических предположений; каждый механизм поддается математическому описанию [43]. Из механизмов мышления концепции-модели наиболее доступны сознанию. Рэй Джакендофф [44] также считает, что наиболее адекватный термин для механизма концепций – это модель, или внутренняя модель мышления. Концепции – это модели в буквальном смысле. Они моделируют в нашем мышлении объекты и ситуации в мире. Например, при зрительном восприятии объекта, концепция-модель в нашей памяти проецирует изображение объекта на зрительную кору мозга. Там это изображение совмещается с изображением, проецируемым из сетчатки глаза.

Концепции используются при удовлетворении основных инстинктов, которые возникли как механизмы выживания задолго до концепций. Гроссберг и Левайн [10] выделили инстинкты, как внутренние сенсоры, измеряющие основные потребности, и отделили их от «инстинктивного поведения», которое надлежит описывать подходящими механизмами. Соответственно, в настоящей статье «инстинкт» – это механизм внутренних сенсоров. Например, когда уровень сахара в крови опускается ниже определённого уровня, инстинкт «сообщает» мозгу, что нужна еда.

Как мы узнаём об инстинктивных нуждах? Мы не слышим инстинктивных повелений и не читаем циферблатов инстинктивных сенсоров. Инстинкты-сенсоры связаны с пониманием и поведением эмоциями. Фундаментальная роль эмоций в системе мышления состоит в том, что эмоциональные нейронные сигналы оценивают концепции для целей удовлетворения инстинктов [10]. Этот эмоциональный механизм оказался принципиально важен для того, чтобы «разорвать замкнутый круг» комбинаторной сложности.

Наш анализ приводит к неотвратимому выводу: человек и высшие животные обладают особым инстинктом, ответственным за познание. Очевидно, человек и высшие животные способны к исследовательскому поведению, даже когда основные телесные потребности, такие как голод, удовлетворены. Биологи и психологи обсуждали различные аспекты этого поведения. Харри Харлоу открыл, что обезьяны и люди стремятся к положительным стимулам, независимо от удовлетворения голода [45]; Дэвид Берлин в этой связи обсуждал любопытство [46]; Леон Фестинджер ввёл понятие когнитивного диссонанса и описал многочисленные эксперименты, раскрывающие потребность человека в уменьшении диссонанса [47]. Фундаментальная природа инстинкта к знанию связана с тем, что наши знания всегда должны быть подогнаны под конкретную окружающую ситуацию. Без адаптации концепций-моделей мы не смогли бы воспринимать непрерывно изменяющийся мир. Мы бы не могли ни ориентироваться, ни удовлетворять наши телесные потребности. Поэтому у нас есть врождённая потребность, инстинкт к улучшению наших знаний. Я назвал его *инстинктом к знанию*. Математически он описывается максимизацией меры близости между концепциями-моделями и миром (как он измеряется сенсорными органами; надо также иметь в виду, что сенсорные органы тоже адаптируются в процессе восприятия).

Эмоции, оценивающие удовлетворение или неудовлетворение инстинкта к знанию, не связаны непосредственно с телесными потребностями. Это, следовательно, духовные или эстетические эмоции. Концептуально-эмоциональное

восприятие мира ведёт к действиям в окружающем мире и к действиям внутри мышления, приводит к увеличению знаний. Математическая теория концептуально-эмоционального восприятия и понимания описывается в следующем разделе. Мышление включает иерархию многих уровней концепций-моделей, от простейших элементов восприятия (линии, движущиеся точки) до концепций-моделей объектов, отношений между объектами, до сложных ситуаций.

## 5. Теория нейронных моделирующих полей (НМП)

Теория нейронных моделирующих полей (НМП) математически описывает механизмы мышления, обсуждавшиеся выше. Это многоуровневая гетеро-иерархическая система [8], это не строгая иерархия, в ней множество обратных связей между соседними уровнями, поэтому это именно гетеро-иерархия. На каждом уровне НМП действуют механизмы концепций-моделей, содержащие знание; они генерируют эфферентные сигналы, распространяющиеся «сверху-вниз» по иерархии и взаимодействующие с афферентными сигналами, поступающими из предыдущих уровней иерархии, «снизу-вверх». Взаимодействие между этими сигналами управляется инстинктом к знанию, который «направляет» концепции-модели к улучшению, то есть направляет обучение, адаптацию, и возникновение новых концепций-моделей для улучшения их соответствия входным сигналам, идущим снизу-вверх.

В этом разделе излагается основной механизм взаимодействия между двумя соседними уровнями иерархии, между афферентными (идущими вверх) и эфферентными (идущими вниз) сигналами. Эти сигналы в нейронной архитектуре представляют собой поля нейронных активностей (этот аспект НМП подобен нейронным активностям в теории адаптивного резонанса [48]), отсюда название этой архитектуры, НМП. Иногда удобнее говорить об этих двух уровнях сигналов, как о входе и выходе одного уровня обработки сигналов. На каждом уровне выходные сигналы – это концепции-модели, воспринятые или образованные из входных сигналов. Входные сигналы ассоциируются с концепциями (или группируются в концепции) в соответствии с моделями и инстинктом к знанию. Эта общая структура НМП соответствует нашим знаниям о нейронной организации мозга; в этой статье, однако, мы не проводим детальное сопоставление математических механизмов с конкретными нейронами и синапсами.

### 5.1. Инстинкт к знанию

Инстинкт к знанию описывается математически как максимизация меры схожести моделей и входных сигналов. В процессе обучения и понимания входных сигналов, модели адаптируются так, чтобы лучше представлять входные сигналы, и чтобы схожесть между ними увеличивалась. Это увеличение схожести удовлетворяет инстинкт к знанию и ощущается как эстетическая эмоция.

На каждом иерархическом уровне нейроны нумеруются индексом  $n = 1, \dots, N$ . Эти нейроны получают входные сигналы  $X(n)$ , идущие снизу-вверх, из нижних уровней иерархии.  $X(n)$  – это поле активности нейронных входных сигналов, поступающих от нейронов на предыдущем (нижнем) уровне. У каждого



нейрона несколько синапсов; в общем случае, мы описываем нейрон набором чисел  $\mathbf{X}(n) = \{X_d(n), d = 1, \dots, D\}$ . Сверху-вниз в эти же нейроны поступают маркирующие активирующие сигналы, идущие от концепций-моделей  $\mathbf{M}_h(\mathbf{S}_h, n)$ ; мы нумеруем модели индексом  $h = 1, \dots, H$ . Каждая модель зависит от параметров этой модели  $\mathbf{S}_h$ . В нейронной структуре мозга параметры кодируются синапсами, силой их связи или весом; математически они задаются набором чисел,  $\mathbf{S}_h = \{S_h^a, a = 1, \dots, A\}$ . Модели являются *представлениями* входных сигналов в следующем смысле. Пусть, например, сигнал  $\mathbf{X}(n)$  поступает от сенсорного нейрона, активированного объектом  $h$ , параметры модели которого  $\mathbf{S}_h$ . Эти параметры могут описывать положение, ориентацию или освещение объекта  $h$ . Модель  $\mathbf{M}_h(\mathbf{S}_h, n)$  *предсказывает* значение входного сигнала  $\mathbf{X}(n)$  в нейроне  $n$ . Например, при зрительном восприятии нейрон  $n$  в зрительной коре получает сигнал  $\mathbf{X}(n)$  из сетчатки и маркирующий сигнал  $\mathbf{M}_h(\mathbf{S}_h, n)$  из концепции-модели объекта  $h$ . Нейрон  $n$  возбуждается, если сильны оба сигнала, идущие снизу-вверх из нижнего уровня и сверху-вниз из модели  $\mathbf{M}_h$ . Ниже описано, как различные модели конкурируют за входные сигналы, в то время как их параметры адаптируются для лучшего соответствия входным сигналам.

Обычное повседневное восприятие использует много уровней от сетчатки до восприятия объекта. В теории НМП одни и те же принципы нейродинамики, действуют на каждом уровне. Восприятие мелких деталей, обычных объектов или понимание сложных абстрактных концепций использует один и тот же механизм, излагаемый ниже. И в восприятии, и в познании участвуют модели и обучение. В восприятии модели соответствуют объектам; в познании – соотношениям и ситуациям.

Обучение состоит в увеличении меры близости между множествами моделей и сигналов  $L(\{\mathbf{X}\}, \{\mathbf{M}\})$ . Мера близости зависит от параметров моделей и от ассоциаций между входными, идущими снизу-вверх сигналами и сигналами от концепций-моделей. Для конкретности мы описываем здесь восприятие объектов упрощённо, как если бы оно происходило в одном уровне.

Сконструируем меру близости  $L(\{\mathbf{X}\}, \{\mathbf{M}\})$ , используя подходы теории вероятностей. При построении выражения для меры близости, необходимо учесть два принципа (почти очевидных). Во-первых, содержание поля зрения неизвестно до восприятия, и, во-вторых, в нем могут присутствовать разнообразные объекты. Важная информация может содержаться в любом из входных сигналов, следовательно, мера близости должна учесть все входные сигналы  $\mathbf{X}(n)$ ,

$$L(\{\mathbf{X}\}, \{\mathbf{M}\}) = \prod_{n \in N} l(\mathbf{X}(n), \{\mathbf{M}\}). \quad (1)$$

Содержащееся здесь произведение по всем входным сигналам удовлетворяет первому принципу: все входные сигналы должны быть учтены (если хотя бы один член произведения равен нулю, мера близости равна нулю, и инстинкт к знанию не удовлетворен). Во-вторых, до восприятия мышлению неизвестно, какой сигнал произошёл от какого объекта. Следовательно, частичная мера близости

сти  $l(\mathbf{X}(n), \{\mathbf{M}\})$  строится так, чтобы каждый потенциальный объект-концепция-модель мог сопоставляться всем возможным входным сигналам. Для этого необходимо рассмотреть все возможные альтернативы по моделям (то есть, провести суммирование по моделям) для каждого входного сигнала. Слагаемые в этой сумме – это условные меры близости между сигналом  $\mathbf{X}(n)$  и моделью  $\mathbf{M}_h$ ,  $l(\mathbf{X}(n)|h)$ . Эта мера «условна»: она предполагает условие, что сигнал  $n$  пришёл от объекта  $h$ , и что этот объект присутствует [49], следовательно, собирая эти слагаемые в общую меру близости,  $L$ , их следует умножить на величину  $r(h)$ , которая представляет вероятностную меру того, что объект  $h$  действительно присутствует. В результате получаем [50]:

$$L(\{\mathbf{X}\}, \{\mathbf{M}\}) = \prod_{n \in N} \sum_{h \in H} r(h) l(\mathbf{X}(n) | h). \quad (2)$$

Структура этого уравнения следует стандартным принципам теории вероятности: сумма берётся по альтернативам  $h$ , а произведение – по различным источникам данных  $n$ . Это выражение не обязательно является вероятностью, но у него вероятностная структура. Если обучение прошло успешно, эта мера близости приближается к вероятностной и ведёт к приблизительно байесовским (то есть оптимальным) решениям. Название «условная мера близости» для  $l(\mathbf{X}(n)|h)$  (или просто  $l(n|h)$ ) следует вероятностной терминологии. Если обучение успешно,  $l(n|h)$  становится условной плотностью вероятности, вероятностной мерой того, что сигнал в нейрон  $n$  пришёл от объекта  $h$ . Тогда  $L$  является полной мерой правдоподобия того, что наблюдаемые сигналы  $\{\mathbf{X}(n)\}$  пришли от объектов, описываемых моделями  $\{\mathbf{M}_h\}$ . Коэффициенты  $r(h)$ , называемые априорными в теории вероятности, содержат информацию о предварительных ожиданиях; их истинные значения неизвестны и подлежат обучению, как и остальные параметры моделей  $\mathbf{S}_h$ . Поскольку  $L$  является мерой правдоподобия, нейронное понятие *обучение* соответствует статистическому понятию оценки в теории оценок. Максимизация правдоподобия обычно ведёт к эффективным или статистически наилучшим оценкам.

Заметим, что в теории вероятности, произведение вероятностей предполагает, что данные независимы. Выражение (2) содержит произведение по  $n$ , однако сигналы  $\mathbf{X}(n)$  не предполагаются независимыми. Зависимость между сигналами существует благодаря моделям; каждая модель  $\mathbf{M}_h(\mathbf{S}_h, n)$  предсказывает ожидаемые значения сигналов для многих нейронов  $n$ . Независимы в выражении (2) лишь отклонения сигналов от модельных предсказаний. То есть, отклонения считаются случайными; неслучайные аспекты сигналов должны быть предсказаны моделями.

Во время процессов обучения-адаптации концепции-модели постоянно изменяются. В настоящей статье мы рассматриваем ситуации, когда функциональные формы моделей  $\mathbf{M}_h(\mathbf{S}_h, n)$ , фиксированы, и обучение-адаптация касается лишь параметров моделей  $\mathbf{S}_h$ . Более сложная структурная адаптация рассмотрена в [51,52]. Время от времени система образует новые модели, иногда старые концепции сливаются или уничтожаются. Это требует изменения меры близости (2); причина состоит в том, что большее число моделей всегда приведет к

лучшей подгонке данных. Это хорошо известная проблема, она решается тем, что меру (2) следует уменьшить, используя «скептическую штрафную функцию»  $p(N, M)$ , которая растёт вместе с числом моделей  $M$ , и скорость её роста возрастает для малого числа данных  $N$ . Например, асимптотически несмещённая оценка по максимуму правдоподобия (говоря простым языком, статистически верная в среднем) требует, чтобы (2) было умножено на  $p(N, M) = \exp(-N_{par}/2)$ , где  $N_{par}$  – это полное число адаптивных параметров во всех моделях (эта штрафная функция известна как информационный критерий Акаика, см. [8] для дальнейших ссылок).

## 5.2. Динамическая логика

Процесс обучения-адаптации состоит в оценке параметров моделей  $\mathbf{S}$  и в ассоциации сигналов с концепциями-моделями; эта оценка производится максимизацией меры (2). Отметим, что (2) содержит все возможные комбинации сигналов и моделей: если раскрыть сумму в (2) и перемножить все члены, то получим  $H^N$  слагаемых. Это огромное число, равное числу комбинаций между всеми входными сигналами ( $N$ ) и всеми моделями ( $H$ ). В этом состоит источник КС прежних алгоритмов. Например, широко известный алгоритм «проверка многих гипотез» пытается максимизировать меру  $L$  по параметрам и по ассоциациям между сигналами и моделями в два приёма. Вначале каждое из  $H^N$  слагаемых (т. е. каждая из ассоциаций сигналы-модели) максимизируется по параметрам. Затем выбирается наибольшее слагаемое, то есть, наилучшая ассоциация для лучшего набора параметров. Ясно, что этот алгоритм сталкивается с проблемой КС, количество вычислений остаётся порядка  $H^N$ .

Теория моделирующих полей решает эту проблему, используя динамическую логику [8,53]. Принципиальный аспект динамической логики состоит в том, что неопределённость или расплывчатость меры близости совпадает с неопределённостью или неточностью моделей. Сначала значения параметров неизвестны и неточность моделей велика; соответственно велика расплывчатость меры близости. В процессе обучения модели становятся более точными и мера близости более чёткой, значение меры  $L$  увеличивается. Приведем математическое описание этого механизма. Сначала зададим произвольные значения неизвестных параметров  $\{\mathbf{S}_h\}$ . Затем вычислим ассоциативные переменные  $f(h|n)$ ,

$$f(h|n) = r(h) l(\mathbf{X}(n)|h) / \sum_{h' \in H} r(h') l(\mathbf{X}(n)|h'). \quad (3)$$

Уравнение (3) подобно байесовской формуле для апостериорных вероятностей; если  $l(n|h)$  в результате обучения станут условными функциями правдоподобия, то  $f(h|n)$  станут байесовскими вероятностями того, что сигнал  $n$  произошёл от объекта  $h$ . Динамическая логика НМП задаётся следующими уравнениями

$$df(h|n)/dt = f(h|n) \sum_{h' \in H} \{[\delta_{hh'} - f(h'|n)] \cdot [\partial \ln l(n|h') / \partial \mathbf{M}_{h'}] \partial \mathbf{M}_{h'} / \partial \mathbf{S}_h\} \cdot d\mathbf{S}_h / dt, \quad (4)$$

$$d\mathbf{S}_h/dt = \sum_{n \in N} f(h|n) [\partial \ln l(n|h) / \partial \mathbf{M}_h] \partial \mathbf{M}_h / \partial \mathbf{S}_h, \quad (5)$$

здесь

$$\delta_{hh'} = 1, \text{ если } h=h', \text{ или } \delta_{hh'} = 0, \text{ если } h \neq h'. \quad (6)$$

Параметр  $t$  – время внутренней динамики системы НМП (число итераций). Для условных мер близости обычно используются гауссовские функции,

$$l(n|h) = G(\mathbf{X}(n) | \mathbf{M}_h(\mathbf{S}_h, n), \mathbf{C}_h). \quad (7)$$

Здесь  $G$  – гауссовская функция со средним  $\mathbf{M}_h$  и матрицей ковариаций  $\mathbf{C}_h$ . Отметим, что в статистике часто используется «гауссовское предположение»; предполагается гауссовская плотность вероятности сигналов. Здесь нет подобного предположения. В (7) сигналы не предполагаются гауссовскими, лишь отклонения сигналов  $\mathbf{X}$  от моделей  $\mathbf{M}$  считаются гауссовскими, так как они возникают в результате многих случайных причин (неслучайные аспекты сигналов учтены в моделях). Если, тем не менее, отклонения не гауссовские, то следует использовать подходящие функции. Если нет информации об их функциональной форме, обычно (7) является хорошим выбором, и не ограничивает общность: взвешенная сумма гауссовских функций в (2) может аппроксимировать любую положительную функцию, каковой является мера близости.

Матрицы ковариаций  $\mathbf{C}_h$  и другие неизвестные параметры в (7) оцениваются эвристически. Первоначальные значения  $\mathbf{C}_h$  должны быть велики, в соответствии с неопределенностью знания моделей  $\mathbf{M}_h$ . По мере улучшения значений параметров и моделей ковариации уменьшаются до истинной величины различий между моделями и сигналами (существующими из-за погрешности измерений и неточности моделей). По мере того, что ковариации уменьшаются, меры близости становятся более четкими, приближаются к дельта-функциям; ассоциативные переменные (3) приближаются к четким значениям  $\{0, 1\}$  и решения динамической логики сходятся к четкой логике (или к вероятностям, в зависимости от задачи). Этот процесс, в котором одновременно улучшаются значения параметров, а мера близости сходится к четкой логической функции, является сутью динамической логики. Таким образом, механизм динамической логики соединяет расплывчатую и четкую логику.

Динамическая эволюция от большой нечеткости к малой – причина для названия «динамическая логика». Математически этот механизм помогает избегать локальных максимумов во время сходимости [8] и соответствует психологическим свойствам [54].

В [8] доказана следующая теорема.

*Теорема.* Уравнения (3)–(6) определяют сходящуюся динамическую систему (НМП) со стационарными состояниями, определяемыми  $\max \{\mathbf{S}_h\} L$ .

Отсюда следует, что стационарные состояния системы НМП являются состояниями максимальной меры близости, удовлетворяющие инстинкт к зна-

нию. Если условные меры близости заданы, как плотности вероятности (или функции правдоподобия), стационарные значения параметров  $\{S_h\}$  являются асимптотически несмещёнными и эффективными оценками [55]. Вычислительная сложность НМП линейна по  $N$ .

Говоря простым языком, это означает, что динамическая логика – это сходящийся процесс. Отдельные аспекты сходимости НМП обсуждаются ниже. Если функции правдоподобия (плотности вероятностей) используются как меры близости, значения параметров оцениваются эффективно (то есть, во многих случаях значения параметров не могут быть улучшены никакой другой процедурой). Приведенная выше теорема также утверждает, что мера близости увеличивается на каждой итерации.

### 5.3 Функционирование динамической логики. Пример

В следующем примере мы ищем «улыбающиеся» и «нахмуренные» модели в шуме (рис. 1). На рис. 1А показаны модели без шума, на рис. 1В – зашумленные модели (сигнал, в котором надо распознать модели). Каждая модель – это 3-х параметрическая параболическая форма (3 параметра описывают положение и кривизну модели-объекта). Размер всего изображения не очень велик, всего 100x100 точек; число объектов равно трем, но оно заранее неизвестно и подлежит определению. Поэтому следует попытаться подогнать к данным по меньшей мере 4 модели, чтобы решить, что 3 модели подходят лучше. Чтобы подогнать  $4 \times 3 = 12$  параметров к 100x100 точкам методом полного перебора, потребовалось бы порядка  $10^{32}$  вычислений. Чтобы применить НМП и динамическую логику для этого примера, следует разработать параметрические модели для ожидаемых объектов, которые в этом случае не слишком сложны [56]: равномерная модель для шума, круглые гауссовские формы для сильно нечётких, плохо различимых объектов и параболические модели для «улыбок» и «нахмурок». Число компьютерных операций в этом примере было около  $10^{10}$ , что значительно меньше, чем при полном переборе.

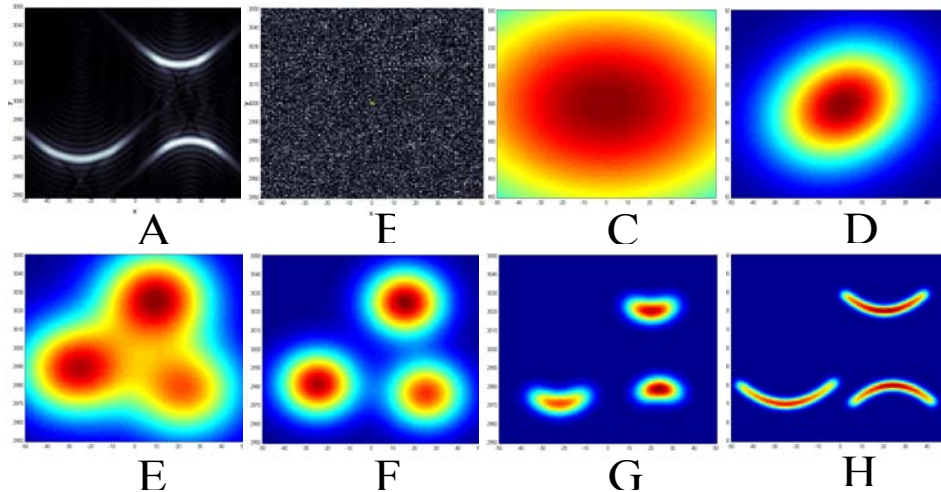


Рис.1 Процесс функционирования динамической логики: нахождение «улыбок» и «нахмурок» в шуме: (A) точная форма «улыбок» и «нахмурок» без шума; (B) реально измеренное изображение с шумом (сигнал слабее шума; отношение сигнала к шуму принимает значения между  $-2\text{dB}$  и  $-0.7\text{dB}$ ; то есть между  $1/3$  и  $1/5$ ); (C) начальная расплывчатая модель; от (D) до (H) показаны улучшающиеся модели на различных стадиях итерационного процесса (всего 22 итерации). После каждых 5 итераций алгоритм пытался увеличить или уменьшить число моделей. Между итерациями (D) и (E) алгоритм пытался подогнать больше одной модели к данным, решил, что ему требуется 3 гауссовские модели для «понимания» данных. Всего было использовано, как уже упоминалось, три типа моделей: одна равномерная модель, описывающая шум и переменное число круглых и параболических моделей. До приблизительно итерации (G) алгоритм «мыслил» в терминах простых круглых моделей, начиная с (G) и далее, алгоритм решил, что для описания данных требуются более сложные параболические модели. Итерации прекратились в (H), когда мера схожести (2) перестала увеличиваться. Этот пример описан подробнее в [56].

Во время процесса адаптации начальные расплывчатые и неопределённые модели ассоциируются со структурами во входных сигналах, нечёткие модели становятся более определёнными и чёткими с каждой итерацией. Тип, форма и число моделей подбираются алгоритмом, в соответствии с уравнениями (3)-(7), так что внутреннее представление внутри системы соответствует входным сигналам: концепции-модели НМП моделируют структуры-объекты в сигналах. На рис. 1A показана точная форма «улыбок» и «нахмурок» без шума; на рис. 1B – реально измеренный сигнал с шумом (отношение сигнала к шуму принимает значения между  $-2\text{dB}$  и  $-0.7\text{dB}$ ; то есть между  $1/3$  и  $1/5$ ); рис. 1C – начальная расплывчатая модель, чем больше расплывчатость, тем больше неопределённость знаний об объектах; от рис. 1D до рис. 1H показаны улучшающиеся модели на различных стадиях итерационного процесса (всего 22 итерации). После каждых 5 итераций алгоритм пытался увеличить или уменьшить число моделей. Между итерациями D и E алгоритм решил, что ему требуется 3 гауссовские модели для «лучшей» подгонки. Всего было использовано, как уже

упоминалось, три типа моделей: одна равномерная модель, описывающая шум, и переменное число круглых и параболических моделей, для которых число моделей, положение и кривизна оцениваются из данных. До приблизительно итерации G алгоритм использовал простые круглые модели, начиная с G и далее, алгоритм решил, что для описания данных требуются более сложные параболические модели. Итерации прекратились в H, когда мера близости перестала увеличиваться.

#### 5.4. Иерархическая организация НМП

Выше мы описали один уровень обработки сигналов в иерархической системе НМП. На каждом уровне иерархии имеются входные сигналы из нижних уровней, модели, меры близости L, эмоции, являющиеся изменениями величины L, и действия; действия внутри мышления включают адаптацию, то есть поведение, удовлетворяющее инстинкт к знанию – максимизацию меры близости в соответствии с уравнениями (3)-(6). Входные сигналы на каждом уровне X(n) представляют собой поля нейронных активностей (возбуждений). В результате обработки сигналов на каждом уровне формируются активированные модели или концепции h, воспринятые или распознанные во входных сигналах n; эти модели вместе с соответствующими инстинктивными сигналами и эмоциями могут активировать модели поведения и вести к действиям на данном уровне.

Активированные модели инициируют и другие действия. Они посылают входные сигналы вверх, на следующий уровень, где создаются формируются более общие, абстрактные модели. Выходные сигналы с данного уровня, служащие входными сигналами на следующем уровне, могут задавать степени активации моделей  $a_h$ , определяемые следующим образом

$$a_h = \sum_{n \in N} f(h|n). \quad (8)$$

Кроме этого, выходные сигналы могут включать параметры моделей. Иерархическая система НМП показана на рис. 2. В иерархии мышления, каждая концепция-модель приобретает «ментальные» смысл и цель на следующем уровне иерархии (в дополнение к другим целям). Например, рассмотрим концепцию модель «стул». «Поведенческая» цель этой концепции – в инициировании действия «сесть» (если наличествует такая телесная потребность), эта «телесная» цель находится на том же иерархическом уровне. Кроме того, у этой концепции есть «чисто ментальная» цель на более высоком уровне иерархии, цель участвовать в распознавании-формировании более общей концепции, скажем «концертный зал», модель которого включает ряды стульев.

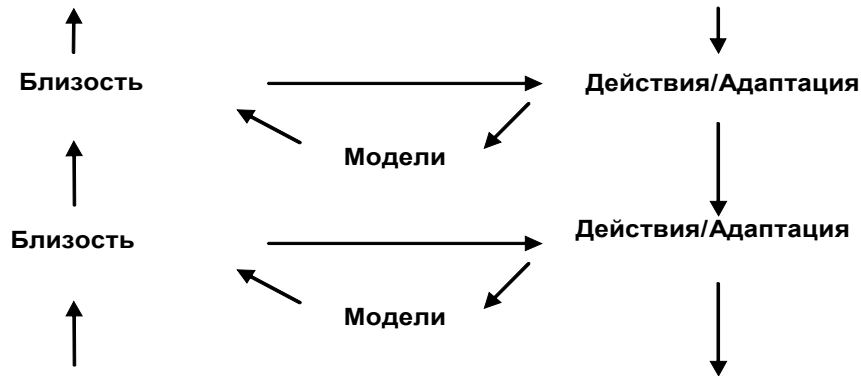


Рис.2. Иерархическая система НМП. На каждом уровне иерархии есть входные сигналы, модели, меры близости, эмоции и действия (в том числе адаптация, максимизирующая меры близости и удовлетворяющая инстинкт к знанию). Большие значения условных мер близости соответствуют концепциям, распознанным на данном уровне, и ведут к активации этих концепций. Сигналы активированных концепций – это выходные сигналы на данном уровне. Они становятся входными сигналами на следующем уровне, распространяя знание вверх по иерархии (на языке физической теории поля, это – пропагаторы концепций-знаний между уровнями иерархии).

Модели на верхних уровнях иерархии более общие и абстрактные, чем модели на нижних уровнях. Например, в зрительной системе, на нижнем уровне иерархии (говоря приблизительно) модели соответствуют ганглиозным нейронам сетчатки и выполняют такие же функции: они распознают простые элементы в поле зрения, линии, движущиеся точки. На более высоких уровнях модели соответствуют функциям, выполняемым в VI-м и в последующих уровнях зрительной коры, там происходит распознавание более сложных зрительных элементов, таких как границы контрастов, их направление, отделение эффектов освещения от геометрических форм, и т.д. Иерархическая структура зрительной коры изучена в деталях [9,57], эти модели могут быть использованы в НМП. В следующих, более высоких познавательных уровнях модели соответствуют объектам, ситуациям, взаимоотношениям между ситуациями, и т.д. [8].  
 Время от времени, как обсуждалось, система НМП формирует новую концепцию или уничтожает (забывает) старую. Более общий механизм НМП, используемый в сложных задачах, работает следующим образом. На каждом уровне система всегда хранит запас неопределённых, расплывчатых, неактивных концепций-моделей (с большими матрицами ковариаций  $C$ ). Эти модели неактивны, поскольку их параметры не адаптированы к сигналам, и поэтому малы их меры близости к сигналам. Такое неактивное состояние самоподдерживается: ассоциативные переменные  $f(h|n)$  для этих моделей малы, в соответствии с уравнением (5) их параметры не адаптируются, и они не участвуют в соревновании между моделями: их меры близости продолжают оставаться малыми для всех сигналов. Однако из-за большой степени расплывчатости (больших ковариаций) их меры близости не обращаются в точности в ноль. Поэтому неактивные модели могут потенциально «ухватить» каждый сигнал, который не попадает в более



специфичные, менее расплывчатые, активные модели. Когда накапливается достаточно сигналов, «выпавших» из активных моделей, и попавших в неактивную модель  $h$ , ее активационный сигнал  $a_h$  превышает определённый порог и модель активируется. Аналогично, когда слишком мало сигналов ассоциировано с определённой моделью, её активационный сигнал уменьшается ниже порогового значения и модель деактивируется. Пороги активации и деактивации обычно определяются на основе информации, существующей на более высоком иерархическом уровне (априорная информация, данные о системных ресурсах, количество активированных моделей различного типа, и т.д.). Активационные сигналы, создаваемые активными моделями на данном уровне  $\{a_h\}$ , формируют «нейронное поле», являющееся входными сигналами для следующего уровня, где формируются более абстрактные и общие концепции.

#### 5.4 Интеграция языка и мышления

Создание интеллектуальных компьютерных систем требует математических методов интегрирования языка и мышления. В НМП это достигается интегрированием языковых и познавательных моделей [51,52]. В уравнениях (1)-(5)  $\mathbf{M}_h$  определяется как двойная модель,

$$\mathbf{M}_h = \{ \mathbf{M}_h^C, \mathbf{M}_h^L \}. \quad (9)$$

Здесь  $\mathbf{M}_h^C$  обозначает познавательную часть модели объекта или ситуации в мире (как модели в примере, представленном на рис. 1);  $\mathbf{M}_h^L$  – это языковая часть модели. Математические механизмы интеграции языка и познания требуют изменений в уравнениях (1)-(5), которые были сформулированы в [58,59]. Эта изменённая система НМП используется сегодня для создания будущих интегрированных систем роботов и «инфоботов», которые, возможно, заменят сегодняшние системы поиска информации на Интернете.

#### 6. Заключение

Итак, в настоящей работе построена модель нейронных модельных полей, позволяющая преодолеть «проклятие размерности» в задачах распознавания и проинтерпретировать ряд особенностей мыслительных процессов.

Уравнения (3)-(6) описывают элементарный процесс восприятия или познания, который максимизирует знание. Знание измеряется мерой близости между концепциями-моделями и миром. В этом процессе большое число концепций-моделей соревнуются за входные сигналы, модели модифицируются, формируются новые; при этом формируются ассоциации между моделями и сигналами, в начале неопределённые, а в результате конкретные, указывающие какой сигнал соответствует какой концепции-объекту-модели. В процессе восприятия входные сигналы поступают от сенсорных органов и соответствуют моделям объектов и простых ситуаций во внешнем мире. В процессе познания

входные сигналы поступают от предыдущих уровней и соответствуют моделям соотношений между концепциями, распознанными на нижних уровнях.

Процессы эти описываются динамической логикой. Её важной математической чертой является соответствие между неопределённостью в моделях и нечёткостью-расплывчатостью ассоциаций  $f(h|n)$ . Пока параметры моделей не соответствуют сигналам, многие модели плохо ассоциированы со многими объектами одновременно. Постепенно одна модель выигрывает соревнование за некий набор сигналов, ассоциации становятся чёткими, и сигналы распознаются как концепции. Начальные нечёткие концепции становятся чёткими, приблизительно подчиняющимися логике.

При подготовке статьи к публикации В.Г. Редько обратил мое внимание на то, что роль эмоций и потребности накопления знаний (аналога инстинкта знаний) при формировании интеллектуального адаптивного поведения подчеркивается в работах А.А. Жданова [60].

## Благодарности

Я с удовольствием выражаю благодарность тем, чьи мысли, идеи, поддержка, и финансирование способствовали написанию этой статьи: М. Арановский, Р. Брокетт, П. Вербос, Б. Виджерс, Д. Виньковецкая, Я. Виньковецкий, Г. Карпентер, Р. Деминг, Ю. Димитрин, В. Дмитриев, Л. Гарвин, И. Гинзбург, Р. Гудвин, С. Гренадер, С. Гроссберг, А. Кангелози, М. Карповский, Л. Заде, Г. Зайниев, С. Кашин, Р. Козма, М. Ксариос, Д. Левайн, Л. Левитин, Р. Линнехан, Т. Лугинбул, М. Мазо, А. Мейстель, С. Миддлтон, К. Мор, К. Мюц, А. Нахман, А. Овсич, Р. Пейн, И. Перловский, В. Петров, Н. Печерская, Ч. Плам, В. Редько, Ю. Румер, Д. Скатруд, Р. Стрейт, Е. Таборская, Т. Терновская, И. Терновский, Ф. Фонтанари, М. Франк-Каменецкий, У. Фрееман, К. Фукунага, А. Черняхов, Е. Шахнович, У. Шёндорф, Дж. Шогрен, Н. Шохирев.

## Литература

1. For the discussions and further references see: Grossberg, S. (1988). *Neural Networks and Natural Intelligence*. MIT Press, Cambridge, MA. Albus, J.S., Meystel, A.M. (2001). *Engineering of Mind: An Introduction to the Science of Intelligent Systems*. Wiley, New York, NY. Meystel, A.M., Albus, J.S. (2001). *Intelligent Systems: Architecture, Design, and Control*. Wiley, New York, NY. Perlovsky, L.I. (2001). *Neural networks and intellect: using model based concepts*. Oxford University Press, NY.
2. *The American Heritage College Dictionary*, Houghton Mifflin, 3rd ed., 2000, Boston, MA. I would emphasize, these general dictionary formulations could only be taken as a starting point.
3. Newell, A. (1983). Intellectual issues in the history of artificial intelligence. In: *The Study of Information*, ed. F.Machlup & U.Mansfield, J.Wiley, New York, NY.
4. Minsky, M. (1988). *The Society of Mind*. MIT Press, Cambridge, MA.
5. Penrose, R. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.

6. Josephson, B. (1997). An Integrated Theory of Nervous System Functioning embracing Nativism and Constructivism. In: *International Complex systems Conference*. Nashua, NH.: Sept 21-26.
7. Hameroff, S. (1994). *Toward a Scientific Basis for Consciousness*. Cambridge, MA.: MIT Press;
8. Perlovsky, L.I. (2001). *Neural Networks and Intellect: Using Model Based Concepts*. New York: Oxford University Press.
9. Grossberg, S. (1988). *Neural Networks and Natural Intelligence*. MIT Press, Cambridge, MA.
10. Grossberg, S., Levine, D.S. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: blocking, inter-stimulus interval, and secondary reinforcement. *Psychobiology*, **15**(3), pp.195-240;
11. Zadeh, L.A. (1997). Information granulation and its centrality in human and machine intelligence. In: *Proceedings of the Conf. on Intelligent Systems and Semiotics '97*. Gaithersburg, MD, pp. 26-30.
12. Meystel, A. (1995). *Semiotic Modeling and Situational Analysis*. AdRem, Bala Cynwyd, PA.
13. G. M. Edelman, G. Tononi, (1995), *A Universe of Consciousness: How Matter Becomes Imagination*, Basic Books, New York, NY.
14. Perlovsky, L.I. & McManus, M.M. (1991). Maximum Likelihood Neural Networks for Sensor Fusion and Adaptive Classification. *Neural Networks*, **4** (1), pp. 89-102.
15. Perlovsky, L.I. (1996). Mathematical Concepts of Intellect. In: *Proceedings, World Congress on Neural Networks*. San Diego, CA.: L. Erlbaum Assoc., pp. 1013-16.
16. Perlovsky, L.I. (2006). Toward physics of the mind: Concepts, emotions, consciousness, and symbols. *Physics of Life Reviews*, **3**(1), pp.22-55.
17. Aristotle, IV BCE, *Complete Works of Aristotle*, Ed.J.Barnes, Princeton, NJ, 1995.
18. Aristotle, IV BCE, *Rhetoric for Alexander*, Complete Works of Aristotle, Ed.J.Barnes, Princeton, NJ, 1995.
19. Aristotle, IV BCE, *Organon*, Complete Works of Aristotle, Ed.J.Barnes, Princeton, NJ, 1995, 18a28-19b4; 1011b24-1012a28.
20. Hilbert, David. (1928/1967). The foundations of mathematics. In J. van Heijenoort, Ed., *From Frege to Gödel*. Cambridge, MA: Harvard University Press, p.475.
21. Gödel, K. (1986). *Kurt Gödel Collected Works*, I. (Ed. S.Feferman et al). Oxford University Press.
22. Об альтернативном истолковании логики см. Gentzen, G. (1935,1936), "Untersuchungen uber das logische Schliessen". *Mathematische Zeitschrift*, 1935, B. 39, pp. 176-210, 405-443. "Die Widerspruchsfreiheit der reinen Zahlentheorie". *Math. Ann.* 1936, Vol. 112, N. 4, pp.493-565. О различных истолкованиях Гёделевской теории и развитии логики после Гёделя см. B. Marchal, 2005, Theoretical Computer Science and the Natural Sciences, *Physics of Life Reviews*, **2**(3), pp.1-38.
23. Perlovsky, L.I. (1998). Conundrum of Combinatorial Complexity. *IEEE Trans. PAMI*, **20**(6) p.666-70.
24. Bellman, R.E. (1961). *Adaptive Control Processes*. Princeton University Press, Princeton, NJ.
25. Minsky, M.L. (1975). A Framework for Representing Knowledge. In: *The Psychology of Computer Vision*, ed. P. H. Winston, McGraw-Hill Book, New York.
26. Winston, P.H. (1984). *Artificial Intelligence*. 2nd edition, Addison-Wesley. Reading, MA.
27. A simple example of an adaptive model is linear regression: the knowledge is encapsulated in using linear model, and in the choice of variables, the uncertainty and adaptivity is in the unknown parameters, fitted to the data. Whereas linear regression uses one model, model-based systems used a large number of models. For example, a

- scene is described using geometric models of many objects. Parameters may include: size, orientation angles, color, illumination conditions, etc. A simple, still nontrivial problem, causing problems in applications till today, is tracking multiple objects in presence of clutter [see ref 29].
28. Singer, R.A., Sea, R.G. and Housewright, R.B. (1974). Derivation and Evaluation of Improved Tracking Filters for Use in Dense Multitarget Environments, *IEEE Transactions on Information Theory*, IT-20, pp. 423-432.
  29. Perlovsky, L.I., Webb, V.H., Bradley, S.R. & Hansen, C.A. (1998). Improved ROTH Detection and Tracking Using MLANS. *AGU Radio Science*, **33**(4), pp.1034-44.
  30. Perlovsky, L.I. (1996). Gödel Theorem and Semiotics. In: *Proceedings of the Conference on Intelligent Systems and Semiotics '96*. Gaithersburg, MD, v.2, pp. 14-18.
  31. Kecman, V. (2001). *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models (Complex Adaptive Systems)*. The MIT Press, Cambridge, MA.
  32. B. Marchal, 2005, Theoretical Computer Science & the Natural Sciences, *Physics of Life Reviews*, **2**(3), pp.1-38.
  33. D.S. Levine, S.J. Leven, Eds. (1992). *Motivation, Emotion, and Goal Direction in Neural Networks*, Erlbaum, Hillsdale, NJ.
  34. Descartes, Rene (1646). *Passions of the Soul*. Descartes deeply penetrated into the interaction between emotions and consciousness: "those who are the most excited by their passions are not those who know them best and their passions are... confused and obscure." He did not differentiate thoughts and emotions consistently: "of all the various kinds of thoughts... these passions," "the action and the passion are thus always one and the same thing." Descartes showed a bias toward unconscious and fused perception of emotions characteristic of a thinking psychological type.
  35. Kant, I. (1790). *Critique of Judgment*, tr. J.H. Bernard, Macmillan & Co., London, 1914.
  36. Sartre J. P., 1948, Emotions. See J. P. Sartre, *Existentialism and Human Emotions*, Citadel Press; Reissue edition, 1984.
  37. Jung, C.G., 1921, Psychological Types. In: *The Collected Works*, v.6, Bollingen Series XX, 1971, Princeton University Press, Princeton, NJ.
  38. Ortony, A, Clore G.L., Collins A., 1990, *The Cognitive Structure of Emotions*, Cambridge University Press
  39. LeDoux, J., 1998, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, Simon & Schuster, New York, NY.
  40. P. E. Griffiths, 1998, *What Emotions Really Are: The Problem of Psychological Categories*, University Of Chicago Press, Chicago, IL.
  41. Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Grosset/Putnam, New York, NY.
  42. Perlovsky, L.I. (1998). *Cyberaesthetics: aesthetics, learning, and control*. STIS'98, Gaithersburg, M.D.; Perlovsky, L.I. (1999). Emotions, Learning, and Control. In: *Proc. International Symp. Intelligent Control, Intelligent Systems and Semiotics*, Cambridge MA, pp.131-137
  43. Perlovsky, L.I. (2006). Modeling Field Theory of Higher Cognitive Functions. In: *Artificial Cognition Systems*, Ed. A. Loula and R. Gudwin, Idea Group, Hershey, PA.
  44. Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford Univ. Press, New York, NY.
  45. Harlow, H.F., Mears, C. (1979). *The Human Model: Primate Perspectives*, Washington, DC: V. H. Winston and Sons.

46. Berlyne, D. E. (1960). *Conflict, Arousal, and Curiosity*, McGraw-Hill, New York, NY; Berlyne, D. E. (1973). *Pleasure, Reward, Preference: Their Nature, Determinants, and Role in Behavior*, Academic Press, New York, NY.
47. Festinger, L. (1957). *A Theory of Cognitive Dissonance*, Stanford, CA: Stanford University Press.
48. Carpenter, G.A., Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine, *Computer Vision, Graphics and Image Processing*, 37, 54-115.
49. Mathematically, the condition that the object  $h$  is present with 100% certainty, is expressed by normalization condition:  $\int l(\mathbf{X} | h) d\mathbf{X} = 1$ . We should also mention another normalization condition:  $\int l(\mathbf{X}(n)) d\mathbf{X}(n) = 1$ , which expresses the fact that each received signal  $n$  is present with 100% certainty. This statement requires a qualification: it does not mean that the system is equally certain about all signals it receives, as discussed in section 5.4, a lower level in the hierarchy is responsible for activating-sensing input signals for a higher level; the lower level “keeps track” of activation levels (certainties) of signals, which it sends up to the higher level.
50. This construction is not unique. Expression (2) is inspired by the notion of likelihood between models and signals. Another general type of similarity measure suitable for the knowledge instinct, inspired the notion of mutual information in the models about the signals, is discussed in [8]. Here we would like to mention a modification of (2) for a specific case. Sometimes a set of observations  $N$  is more convenient to describe mathematically as a continuous flow of signals, for example, a flow of visual stimuli in time and space; then, it is convenient instead of eq.(1) to consider its continuous version,

$$L = \exp \int_N \ln \left( \sum_{h \in H} r(h) l(\mathbf{X}(n) | h) \right),$$

where  $N$  is a continuum of bottom-up signals, such as in time-space.

51. Perlovsky, L.I. (2004). Integrating Language and Cognition. *IEEE Connections*, Feature Article, 2(2), pp. 8-12.
52. Perlovsky, L.I. (2006). Symbols: Integrated Cognition and Language. In: *Computational Semiotics*, Ed. A. Loula and R. Gudwin, Idea Publishing, Hershey, PA.
53. Perlovsky, L.I. (1996). Mathematical Concepts of Intellect. In: *Proc. World Congress on Neural Networks*, San Diego, CA; Lawrence Erlbaum Associates, NJ, pp.1013-16;
- Perlovsky, L.I.(1997). Physical Concepts of Intellect. *Proc. Russian Academy of Sciences*, 354(3), pp. 320-323.
54. A desirable property of convergence of model parameters to their correct values for objects they represent, or to the global maximum of the likelihood is not necessary in practical computer applications or in workings of the mind all the time. Perceptions are not always correct at a first sight, sometimes, we have to move, to change a point of view, before we unambiguously recognize an object. Same is true in computer applications. The practical aspects of efficient convergence are considered in [8, 29].
55. Cramer, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton NJ.
56. Linnehan, R., Mutz, Perlovsky, L.I., C., Weijers, B., Schindler, J., Brockett, R. (2003). Detection of Patterns below Clutter in Images. In: *Int. Conf. on Integration of Knowledge Intensive Multi-Agent Systems*, Cambridge, MA Oct.1-3, 2003.
57. Zeki, S. (1993). *A Vision of the Brain* Blackwell, Oxford, England.
58. Perlovsky, L.I. (2004). Integrating Language and Cognition. *IEEE Connections*, Feature Article, 2(2), pp. 8-12.

Л.И. Перловский

59. Perlovsky, L.I. (2006). Neural Networks, Fuzzy Models and Dynamic Logic. Chapter in R. Köhler and A. Mehler, eds., *Aspects of Automatic Text Analysis* (Festschrift in Honor of Burghard Rieger), Springer, Germany, pp.363-386.
60. Жданов А.А. Бионический метод автономного адаптивного управления // *От моделей поведения к искусственному интеллекту* (под ред. Редько В.Г.). М.: УРСС. Серия "Науки об искусственном". 2006. С. 343-385.